Maximum entropy Edgeworth estimates of the number of integer points in polytopes

Alexander Barvinok and J.A.Hartigan

3 August 2010

Abstract: *The number of points $x = (x_1, x_2, ...x_n)$ that lie in an integer cube C in R^n and satisfy the constraints $\sum_j h_{ij}(x_j) = s_i, 1 \le i \le d$ is approximated by an Edgeworth-corrected Gaussian formula based on the maximum entropy density p on $x \in C$, that satisfies $E \sum_j h_{ij}(x_j) = s_i, 1 \le i \le d$. Under p, the variables $X_1, X_2, ...X_n$ are independent with densities of exponential form. Letting S_i denote the random variable $\sum_j h_{ij}(X_j)$, conditional on $S = s, X$ is uniformly distributed over the integers in C that satisfy $S = s$. The number of points in C satisfying $S = s$ is $p\{S = s\} \exp(I(p))$ where $I(p)$ is the entropy of the density p. We estimate $p\{S = s\}$ by $p_Z(s)$, the density at s of the multivariate Gaussian Z with the same first two moments as S; and when d is large we use in addition an Edgeworth factor that requires the first four moments of S under p. The asymptotic validity of the Edgeworth-corrected estimate is proved and demonstrated for counting contingency tables with given row and column sums as the number of rows and columns approaches infinity, and demonstrated for counting the number of graphs with a given degree sequence, as the number of vertices approaches infinity.*

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043, USA E-mail address: barvinok@umich.edu

Department of Statistics, Yale University, New Haven, CT 06520-8290

1 Maximum entropy estimation of the number of integer points

Let $x = (x_1, x_2, ... x_n)$ be a vector in R^n. For arbitrary $R^1 \to R^1$ functions h_{ij} define $S_i = \sum_j h_{ij}(x_j), 1 \leq i \leq d$. Let Q be counting measure on a cube C of integers in R^n. Consider the surface $S = s$ in R^n consisting of points x that satisfy the sums $S_i = \sum_j h_{ij}(x_j) = s_i, 1 \leq i \leq d$. The volume of the surface $Q\{S = s\} > 0$ is the number of points that lie in C and in the surface $\{S = s\}$. For P_U the uniform distribution on the cube C, $Q\{S = s\} = P_U\{S = s\}Q\{C\}$.

Let $X = (X_1, X_2, ... X_n)$ be n random variables uniformly distributed over the cube. Since the random variables are independent, the central limit theorem will apply to the sums $S_i = \sum_j h_{ij}(X_j)$ under suitable conditions on the h. Thus we might approximate the probability $P_U\{S = s\}$ by $p_Z(s)$, the density at s of a multivariate Gaussian Z with the same first and second moments as S. We expect this approximation to work well when the mean of S is close to the selected values s, but not so well in the tails of the distribution. Therefore we propose maximum entropy Gaussian estimation of the volume using an approximating Gaussian with mean value s. This procedure is called *exponential tilting*; see, for example, [KT03].

The entropy of a discrete random variable X having density p (with respect to counting measure) is:

(1) $$I(p) = -E\{\log p(X)\}.$$

We find the maximum entropy distribution P described in [J57] , with density p on a cube C of integers in R^n satisfying $ES = s$. If there is a density of exponential form

(2) $$P\{X = x\} = p(x) = \exp\{\sum_{ij} \lambda_i h_{ij}(X_j) + \lambda_0\}$$

where the λ_i are chosen to satisfy the expectations $ES = s$, and to ensure that $\sum_{x \in C} p(x) = 1$, then this density may be shown to be the unique maximum entropy density subject to the constraints $ES = s$.

Under P, the variables $X_1, X_2, ... X_n$ are independent with densities

(3) $$p_j(x_j) = \exp\{\sum_i \lambda_i h_{ij}(x_j) + \upsilon_j\}.$$

And, conditional on $S = s$, X is uniformly distributed over the integers x in C that satisfy $S = s$, with

(4) $$p(x) = \exp\{\sum_i \lambda_i s_i + \lambda_0\} = \exp\{-I(p)\},$$

since

(5) $$I(p) = \sum_x [-p(x) \log p(x)] = -E\{\sum_{ij} \lambda_i h_{ij}(X_j) + \lambda_0\} = -\sum_i \lambda_i s_i - \lambda_0$$

Thus, for any x that satisfies $S = s$,

(6) $$Q\{S = s\} = P\{S = s\}/p(x) = P\{S = s\} \exp\{I(p)\}.$$

The entropy term in this formula was suggested in some special cases in [B09].

We again estimate $P\{S = s\}$ by $p_Z(s)$, the density at s of a multivariate Gaussian

Z with the mean and covariance of S. The advantage in using the maximum entropy P is that the mean of the Gaussian is s, so that "debiased" estimation takes place at the mean.

If the h functions are just multiples, say $h_{ij}(X_j) = A_{ij}X_j$, then the maximum entropy density p consists of independent exponential form densities

$$(7) \qquad p_j(x) = \exp\{\theta_j x - c(\theta_j)\}$$

on the X_j with canonical parameters $\theta_j = \sum_i \lambda_i A_{ij}$ and expectations $c'(\theta_j)$. The parameters λ_i are chosen so that $\sum_j A_{ij} c'(\theta_j) = s_i$.

Because p is maximum entropy, the θ_j may also be characterized [Ba09] as the unique maxima of the p-entropy $\sum_j [c(\theta_j) - \theta_j c'(\theta_j)]$ for a given θ subject to $\sum_j A_{ij} c'(\theta_j) = s_i$. And then

$$(8) \qquad Q\{S = s\} = P\{S = s\}/\prod_j \exp\{\theta_j c'(\theta_j) - c(\theta_j)\} = P\{S = s\} \exp\{I(p)\}.$$

So far, we have followed the approach in [BH10a] of maximum entropy gaussian approximation. However, when the number d of sums S_i approaches infinity, and the variances of the sums are $O(d)$, the relative error in Gaussian approximation to the true density for the ith sum will be typically $P\{S_i = s_i\}/p_{Z_i}(s_i) - 1 = O(1/d)$ and the error in approximating the true density for d sums will be about $(1 + O(1/d))^d - 1 = O(1)$. In order to get an accurate approximation we need to consider the Edgeworth corrections to the Gaussian approximation, which use the third and fourth cumulants of the S distribution.

In [MW90], McKay and Wormald produced an asymptotic formula for the number of near regular graphs on n vertices with k edges, where k is proportional to n. They derive the formula by a saddlepoint approximation to Cauchy's integral for determining a coefficient in a generating function. Their generating function turns out to be the characteristic function of the sums S appropriate for this problem. The maximum entropy Edgeworth approximation generalises their formula to graphs with widely varying degree sequences in [BH10b]. The maximum entropy method can also be used to estimate the number of graphs with given degree sequences and with additional edge specifications such as specified cliques or colorings of the graph.

In [CM05], [GMW06], [CM07], [CGM08], [MG] ,Canfield, Greenhill, McKay, Wormald, and Wang extended the Cauchy integral approach to asymptotic enumeration of two way contingency tables of integers in which the marginal sums are known, with the row sums nearly equal and the column sums nearly equal. The integers may be non-negative, or constrained to be 0-1. The maximum entropy Edgeworth approximation, (see also [BH09]), generalises their formulae to the case of varying marginal sums. The formulae require the first four moments of certain sums of independent random variables. The maximum entropy table entries are independent geometric variables when the integers in the tables are non-negative, and independent Bernoulli variables when the integers are 0-1.

The advance in the maximum entropy Edgeworth approximation is that it provides a unified method for the problems mentioned above, and for generalisations of them, using a standard statistical approximation,(see for example [K06]), based

on the first four moments of sums of independent variables determined by the maximum entropy distributions.

Diaconis and Efron [DE85] study the distribution of a chi-square statistic for the uniform distribution over contingency tables with fixed margins. The number of rows and columns are fixed, but the total count approaches infinity. If instead the table entries are bounded, but the numbers of rows and columns approach infinity, we expect that a maximum entropy approach should yield a valid asymptotic estimate of the distribution. Here the maximum entropy table entries are integer Gaussians: Gaussian variables, with arbitrary means and variances, constrained to be integers.

2 The Edgeworth approximation for integer random variables of increasing dimensionality.

Let X_d be a sequence of $d-$dimensional integer random variables having mean 0. Suppose that the determinant of the lattice generated by values of X_d having positive probability is Δ_d. We wish to estimate the probability $P\{X_d = 0\}$ using the first four moments of X_d.

Define $Q_a(t) = 1$ if $\max_i |t_i| \leq a, Q_a(t) = 0$ if $\max |t_i| > a$.

We use the d-dimensioned characteristic function $\phi_d(t) = E(it'X_d)$, with t a column vector in R^d, and t' the corresponding row vector:

$$(9) \qquad P\{X_d = 0\} = (2\pi)^{-d} \int Q_\pi \phi_d(t).$$

The cumulant term $K_d^r(t)$ is the polynomial term of degree r in the expansion $\log \phi_d(t) = \sum_{r=1}^\infty \frac{i^r}{r!} K_d^r(t)$. Specifically,

$$(10) \quad K_d^2(t) = E(t'X_d)^2, \quad K_d^3(t) = E(t'X_d)^3, \quad K_d^4(t) = E(t'X_d)^4 - 3(K_d^2(t))^2.$$

The variance-covariance matrix V_d, is determined by the second cumulant:

$$(11) \qquad \sum_{ij} t_i t_j V_d(i,j) = K_d^2(t).$$

Define $\kappa_d^3 = E_d\{K_d^3(t)^2\}, \kappa_d^4 = E_d K_d^4(t)$ where the expectation E_d is with respect to $t \sim N(0, V_d^{-1})$, a Gaussian variable with mean 0 and variance-covariance V_d^{-1}. The Edgeworth approximation to $P\{X_d = 0\}$ is

$$(12) \qquad \hat{P}\{X_d = 0\} = \Delta_d(2\pi)^{-d/2}|V_d|^{-1/2}\exp(-\kappa_d^3/72 + \kappa_d^4/24).$$

The approximation consists of the density at zero of a Gaussian with variance-covariance V_d, multiplied by an Edgeworth term correcting for the departure from Gaussianity.

We will use the order of magnitude notation

$$(13) \qquad f(d) = o(g(d)) \quad : \quad f(d)/g(d) \to 0 \text{ as } d \to \infty,$$

$$(14) \qquad f(d) = O(g(d)) \quad : \quad \limsup_d |\frac{f(d)}{g(d)}| < \infty.$$

Theorem 1 Let E_d denote expectation with respect to $t \sim N(0, V_d^{-1})$. Suppose that for some M, $\quad \varepsilon = M\sqrt{\log d/d}$,

$$(I) \qquad \kappa_d^3 = O(1), \quad \kappa_d^4 = O(1),$$

$$(II) \qquad E_d\left\{Q_\varepsilon \exp[\frac{1}{12}K_d^4]\right\} = O(1),$$

$$(III) \qquad Q_\varepsilon[\log\phi_d(t) - \sum_{r=2}^4 K_d^r(t)\frac{i^r}{r!}] \quad = \quad o(1),$$

$$(IV) \quad E_d\left\{Q_\varepsilon \exp[-\frac{1}{6}iK_d^3(t) + \frac{1}{72}\kappa_d^3 + \frac{1}{24}K_d^4(t) - \frac{1}{24}\kappa_d^4]\right\} \quad \to \quad 1 \text{ as } d \to \infty,$$

$$(V) \qquad \int_{Q_\pi - Q_\varepsilon} |\phi_d(t)|/\int_{Q_\varepsilon} |\phi_d(t)| \quad = \quad o(1).$$

(13) $\qquad$ Then $P\{X_d = 0\}/\hat{P}\{X_d = 0\} \to 1$ as $d \to \infty$.

Comments on conditions:

The theorem doesn't prove too much itself, but rather outlines a program for proving the validity of the approximation in particular cases.

Conditions I,II bound the third and fourth cumulants. Condition III,IV require that the third and fourth cumulants affect the characteristic function integral through the summary cumulants κ_d^3, κ_d^4. Condition V requires that contributions to the characteristic function integral be negligible outside a small cube centered at 0. In particular this causes the determinant of the lattice of possible values of X_d to be 1 for d large enough.

Proof: Let $K_{34}(t) = -\frac{1}{6}iK_d^3(t) + \frac{1}{72}\kappa_d^3 + \frac{1}{24}K_d^4(t) - \frac{1}{24}\kappa_d^4$.

From I,II,

$$
\begin{aligned}
E_d\{Q_\varepsilon|\exp K_{34}(t)|^2\} &\leq \exp[\tfrac{1}{36}\kappa_d^3 + \tfrac{1}{12}K_d^4(t) - \tfrac{1}{12}\kappa_d^4] = O(1), \\
E_d\{Q_\varepsilon \exp[K_{34}(t) + o(1)]\} &- E_d\{Q_\varepsilon \exp[K_{34}(t)]\} \\
&= o(1)E_d\{Q_\varepsilon \exp[K_{34}(t)]\} \\
&= o(1)\left(E_d\{Q_\varepsilon|\exp K_{34}(t)|^2\}\right)^{1/2} = o(1),
\end{aligned}
$$

(14)

From III,IV

(15)
$$
\begin{aligned}
\Delta_d(2\pi)^{-d}\int Q_\varepsilon(t)\phi_d(t)/\hat{P}(X_d = 0) &= E_d\{Q_\varepsilon\exp[\tfrac{1}{2}K_d^2(t) + \log\phi(t) + \tfrac{1}{72}\kappa_d^3 - \tfrac{1}{24}\kappa_d^4]\} \\
&= E_d\{Q_\varepsilon\exp[K_{34}(t) + o(1)]\} = E_d\{Q_\varepsilon\exp[K_{34}(t)]\} + o(1) \to 1 \text{ as } t \to \infty.
\end{aligned}
$$

Thus

(16) $\qquad \displaystyle\int Q_\varepsilon\phi_d/\left\{(2\pi)^{d/2}|V_d|^{-1/2}\exp[-\frac{1}{72}\kappa_d^3 + \frac{1}{24}\kappa_d^4]\right\} \to 1.$

A similar argument shows that, since $|\exp[K_d^3(t)i^3/3!]| = 1$,

(17) $\qquad \displaystyle\int Q_\varepsilon|\phi_d|/\left\{(2\pi)^{-d/2}|V_d|^{1/2}\exp[\frac{1}{24}\kappa_d^4]\right\} \to 1.$

This shows that $\int Q_\varepsilon|\phi_d| = O(1)|\int Q_\varepsilon\phi_d|$.

Thus from condition V ,

(18) $\qquad \displaystyle\int Q_\pi\phi_d/\int Q_\varepsilon\phi_d \to 1.$

We now show that condition V requires the determinant Δ_d of the lattice to be 1 for d large enough. In the contrary case, consider the reciprocal lattice in d dimensions consisting of all vectors a for which $a'X_d$ is integer with probability one. The determinant of this lattice is the reciprocal of the determinant of the original lattice, and so the reciprocal determinant is less than or equal to $^1/_2$. There must be a non-zero point in the reciprocal lattice which lies in the half-unit cube; thus there is a non-zero point $t = 2\pi a$ lying in the cube $Q_\pi(t) = 1$ for which $a'X_d$ is integer. Now $\phi_d(t + u) = E\{\exp(i(t + u)'X_d)\} = E\{\exp(iu'X_d)\} = \phi_d(u)$, since $\exp(2\pi a'X_d) = 1$. Thus the integral $|\phi_d(t)|$ in the neighbourhood of $t = 2\pi a$ equals its integral in the neighbourhood of 0, which contradicts V.

Since $\Delta_d = 1$, combining (15) and (18) gives

$$(19) \qquad P\{X_d = 0\}/\hat{P}\{X_d = 0\} = (1 + o(1)) \int Q_\pi \phi_d / \int Q_\varepsilon \phi_d \to 1 \text{ as } d \to \infty,$$

which concludes the proof.

3 Numbers of contingency tables with given row and column sums

Consider a contingency table of non-negative integers $X_{ij}, 1 \le i \le m, 1 \le j \le n$ with row and column sums $R_i = \sum_j X_{ij}, C_j = \sum_i X_{ij}$. We wish to estimate the number of tables satisfying the constraints $R_i = r_i, C_j = c_j$. Define the dimension $d = (m + n - 1)$ integer vector S_d :

$$(20) \qquad \begin{aligned} S_{jd} &= R_j - r_j, 1 \le j \le m, \\ S_{(k+m)d} &= C_k - c_k, 1 \le k \le n - 1. \end{aligned}$$

Following the program of section 2, the Edgeworth approximation begins with the maximum entropy distribution for $\{X_{jk}\}$ with expectations $ER_j = r_j, EC_k = c_k$, which consists of independent geometrics with expectations μ_{jk}:

$$(21) \qquad P\{X_{jk} = x\} = (\frac{\mu_{jk}}{1 + \mu_{jk}})^x / (1 + \mu_{jk}),$$

where $\log(1 + 1/\mu_{jk}) = \alpha_j + \beta_k$ and parameters α_j, β_k are chosen for which

$$(22) \qquad ER_j = \sum_k \mu_{jk} = r_j, EC_j = \sum_j \mu_{jk} = c_k.$$

The existence of parameters α_j, β_k satisfying the marginal constraints is shown in [B09]. The maximum entropy entries μ_{jk} are uniquely determined. $\alpha + c, \beta - c$ is a solution if and only if α, β is a solution.

The conditional distribution of $\{X_{jk}\}$ under the constraints $\{R = r, C = c\}$, (equivalently $\{S_d = 0\}$), is uniform. The number of integers satisfying the constraints is

$$(23) \quad Q(S_d = 0) = P\{S_d = 0\} \exp(I(P)) = P\{S_d = 0\} \prod_{jk} (1 + \mu_{jk})^{1 + \mu_{jk}} \mu_{jk}^{-\mu_{jk}}.$$

The probability $P\{S_d = 0\}$ is approximated by

$$(24) \qquad \hat{P}\{S_d = 0\} = (2\pi)^{-d/2} |V_d|^{-1/2} \exp(-\kappa_d^3/72 + \kappa_d^4/24),$$

depending on the first four cumulants of S_d, as explained in section 2. See [BH09,BH10a] for further discussion.

Each element of S_d is the deviation from its mean of a sum of independent geometrics with expectations $\{\mu_{jk}\}$. The mean-centered geometric characteristic function with expectation μ is

$$(25) \qquad \psi_\mu(t) = e^{-i\mu t} / (1 - \mu(e^{it} - 1)).$$

From theorem 1, the validity of the asymptotic estimate may be assessed by the limiting behavior of the characteristic function of S_d, with parameters

$$(26) \qquad \begin{aligned} & t_j = v_j, 1 \le j \le m, \quad t_{m+k} = w_k, 1 \le k \le n - 1, \quad w_n = 0, \\ \phi_d(t) \quad &= \quad E\left\{\exp\left(i[v'(R-r) + w'(C-c)]\right)\right\} = \prod_{jk} \psi_{\mu_{jk}}(v_j + w_k) \end{aligned}$$

We sometimes use t to refer to all the parameters in the characteristic function, and at other times use v, w to treat separately the parameters in the characteristic function associated with the rows and columns respectively.

Use $x_n \sim y_n$ if $x_n/y_n \to 1$ and $x_n \approx y_n$ if $\limsup |x_n/y_n| < \infty, \limsup |y_n/x_n| < \infty$.

Theorem 2

Suppose that , as $d = m + n - 1 \to \infty, m \approx n \approx \min r_i \approx \max r_i \approx \min c_i \approx \max c_i$. Assume that

$$(27) \qquad \liminf \frac{(1 + \frac{n}{\max r_i})(1 + \frac{m}{\max c_i})}{(1 + \frac{mn}{\sum r_i})} > 1,$$

The cumulants $K_d^r(t)$ of $t'S_d$ are the sums of the corresponding cumulants of the geometrics with expectations μ_{jk} and parameters $t_{jk} = v_j + w_k$,

$$(28) \qquad \begin{aligned} K_d^2 &= \sum_{jk} t_{jk}^2 \mu_{jk}(1 + \mu_{jk}) \\ K_d^3 &= \sum_{jk} t_{jk}^3 \mu_{jk}(1 + \mu_{jk})(1 + 2\mu_{jk}) \\ K_d^4 &= \sum_{jk} t_{jk}^4 \mu_{jk}(1 + \mu_{jk})(1 + 6\mu_{jk}(1 + \mu_{jk})). \end{aligned}$$

Let $K_d^2 = t'V_d t$. Let E_d denote expectation with respect to $t \sim N(0, V_d^{-1})$. Then

$$(29) \qquad P\{S_d = 0\}/(2\pi)^{-d/2}|V_d|^{-1/2} \exp(-\frac{1}{72}E_d\{(K_d^3)^2\} + \frac{1}{24}E_d K_d^4)\} \to 1$$

Remark on conditions: Our proof requires that the relative sizes of the maximum entropy entries be bounded asymptotically, and that the absolute sizes are bounded away from zero and infinity. In [BH09] we prove validity of the Edgeworth approximation dropping the condition that the absolute sizes be bounded away from infinity.

Proof:

We will show conditions I-V of theorem 1 hold.

Lemma 3.1 $\max \mu_{ij} \approx \min \mu_{ij} \approx 1$.

Proof:

Let Ω be the set of $m \times n$ matrices μ satisfying

$$(30) \qquad \begin{aligned} \mu_{ij} &> 0, \\ i \leq k, j \leq l \quad \Rightarrow \quad \mu_{ij} &\geq \mu_{kl}, \\ (1 + 1/\mu_{ij})(1 + 1/\mu_{kl}) &= (1 + 1/\mu_{il})(1 + 1/\mu_{kj}). \end{aligned}$$

Since the previous equation holds if and only if $\log(1 + 1/\mu_{jk}) = \alpha_j + \beta_k$, these matrices consist of the maximum entropy geometric expectation matrices corresponding to the possible non-increasing positive row sums $r_1 \geq r_2 \geq ... r_m > 0$ and the possible non-increasing positive column sums $c_1 \geq c_2 \geq ... c_n > 0$.

Lemma 3.1.1

The maximum entry μ_{11} achieves its maximum over $\mu \in \Omega$ for given values of $r_1 = \sum_k \mu_{1k}, c_1 = \sum_j \mu_{j1}, T = \sum_{jk} \mu_{jk}$ when $\mu_{12} = \cdots = \mu_{1j} = \cdots = \mu_{1n}, \mu_{21} = ... \mu_{i1} \cdots = \mu_{m1}$. And the minimum entry μ_{mn} achieves its minimum for given values of r_m, c_n, T when $\mu_{m1} = ... \mu_{1j} \cdots = \mu_{m(n-1)}, \mu_{1n} = ... \mu_{2n} \cdots = \mu_{(m-1)n}$.

Proof:

The result is trivial if either $r_1 = T/m$ or $c_1 = T/n$; it will be useful, for uniqueness, to forbid these conditions.

We first prove that the maximum entry μ_{11} achieves its maximum over $\mu \in \Omega$ for given values of $r_1 = \sum_k \mu_{1k}, c_1 = \sum_j \mu_{j1}, T = \sum_{jk} \mu_{jk}$ when $\mu_{12} = \cdots = \mu_{1j} = \cdots = \mu_{1n}, \mu_{21} = ... \mu_{i1} \cdots = \mu_{m1}$. Equivalently, since by (30), μ is determined by its first row and column, it is equivalent to maximize μ_{11} over choices of $u = \{\mu_{j1}, 2 \leq j \leq m\}, v = \{\mu_{1k}, 2 \leq k \leq n\}$, for given values of r_1, c_1, T. We need to

show that the maximal μ occurs when $(u, z) \in \Xi$, where all the $u's$ are equal and all the $z's$ are equal.

Consider first the maximization of T over $\mu \in \Omega$ with r_1, c_1, μ_{11} fixed, which is equivalent to maximizing T over choices of (u, z) which are constrained to lie in a compact polyhedron so that r_1, c_1, μ_{11} are fixed.

Add a further constraint by fixing z, so that the maximization occurs by varying only the entries u. From (30), for $i > 1$,

$$(31) \qquad (1 + 1/\mu_{ij}) = (1 + 1/\mu_{i1})(1 + 1/\mu_{1j})/(1 + 1/\mu_{11}) = \lambda_j(1 + 1/\mu_{i1})$$

where $\lambda_j = (1 + 1/\mu_{1j})/(1 + 1/\mu_{11}) \geq 1$ is fixed given the first row, and by the forbidden equality, $\lambda_j > 1$ for at least one j. For $i > 1$, it follows that μ_{ij} is a concave function of μ_{i1} determined by the fixed λ_j, and by the forbidden equality $\sum_j \mu_{ij} = g(\mu_{i1})$ where g is strictly concave in μ_{i1}, and depends only on the fixed λ_j.

Thus $T = \sum_j \mu_{1j} + \sum_{i>1} g(\mu_{i1})$ is a strictly concave function of u with a unique maximum at u^0, say. If $u_{i1}^0 \neq u_{(i+1)1}^0$, then by strict concavity of g, $2g(\frac{1}{2}[u_{i1}^0 + u_{(i+1)1}^0]) > g(u_{i1}^0) + g(u_{(i+1)1}^0)$, so the function T may be improved by replacing both u_{i1}^0 and $u_{(i+1)1}^0$ by $\frac{1}{2}[u_{i1}^0 + u_{(i+1)1}^0]$, a contradiction. Thus $\mu_{21} = \cdots = \mu_{i1} = \cdots = \mu_{m1}$ at the maximum.

Now return to the maximization of T over u, z with r_1, c_1, μ_{11} fixed. The maximum of T, say $T(\mu_{11})$, occurs for some (u, z), and it may be improved, from the previous paragraph, unless $(u, z) \in \Xi$, so these conditions hold at the maximum. In addition, the maximizing point (u, z) is unique, given r_1, c_1, T. Thus, at the maximum,

$$(32) \qquad \begin{aligned} T &= r_1 + c_1 - \mu_{11} + (m-1)(n-1)\mu_{22} \\ 1 + \frac{1}{\mu_{22}} &= (1 + \frac{m-1}{c_1-\mu_{11}})(1 + \frac{n-1}{r_1-\mu_{11}})\frac{\mu_{11}}{1+\mu_{11}} \end{aligned}$$

It will be seen from (32) that μ_{22} and therefore $T(\mu_{11})$ are both decreasing functions of μ_{11}.

Finally, we turn to the maximization of μ_{11} over $\mu \in \Omega$ with $r_1, c_1, T = T^0$ fixed, accomplished by considering all choices of u, z constrained to lie in a compact set Γ so that $r_1, c_1, T = T^0$ are fixed. Then $\mu_{11} = \mu_{11}^0$ is maximized at some point (u^0, z^0) in Γ. If $(u^0, z^0) \notin \Xi$, we can find, $(u^1, z^1) \in \Xi$ maximizing T for the given r_1, c_1, μ_{11}^0, so that $T(\mu_{11}^0) > T^0$. Since μ_{11}^0 is maximal, the value of μ_{11} at the point (u^1, z^1) given $r_1, c_1, T = T^0$ must satisfy $\mu_{11}^1 \leq \mu_{11}^0$. Also, the maximal value of T given r_1, c_1, μ_{11}^1 is achieved at the unique point $(u^1, z^1) \in \Xi$, so that $T^0 = T(\mu_{11}^1)$. Since $T(\mu_{11})$ is decreasing in μ_{11}, $T^0 = T(\mu_{11}^1) \geq T(\mu_{11}^0)$ which contradicts $T(\mu_{11}^0) > T^0$ and establishes that the maximum of μ_{11} over $\mu \in \Omega$ with $r_1, c_1, T = T^0$ fixed occurs at a point $(u, z) \in \Xi$.

A similar argument is used for the minimum of μ_{11}, first minimizing T for fixed r_m, c_n, μ_{mn} over possible choices of the last row and column, showing that that the values of last column other than the last entry are equal, and the values of the last row other than the last entry are equal. And then transfer this result to the minimization of μ_{mn} for fixed r_m, c_n, T.

This concludes the proof of lemma 3.1.1.

From lemma 3.1.1, the minimum entry μ_{mn} for given r_m, c_n, T occurs when $\mu_{m1} = \cdots = \mu_{1j} = \cdots = \mu_{m(n-1)}, \mu_{1n} = \cdots = \mu_{2n} = \cdots = \mu_{(m-1)n}$. In this case

(33)
$$
\begin{aligned}
(n-1)\mu_{m1} + \mu_{mn} = r_m \Rightarrow \mu_{m1} &= O(1), \\
(m-1)\mu_{1n} + \mu_{mn} = c_n \Rightarrow \mu_{1n} &= O(1), \\
(n-1)\mu_{11} + \mu_{1n} = r_1 \Rightarrow \mu_{11} &= O(1), \\
1 + 1/\mu_{mn} = (1 + 1/\mu_{m1})(1 + 1/\mu_{1n})/(1 + 1/\mu_{11}) &= O(1).
\end{aligned}
$$

This guarantees that μ_{mn} is bounded away from zero in the extreme case where it takes its smallest value, so it must be bounded away from zero in every case. Also $\mu_{mn} < r_1/n$ is bounded away from ∞ by the first assumption. Thus $\mu_{mn} \approx 1$ as required.

The maximum entry μ_{11} for given m, n, r_1, c_1, T occurs when $\mu_{12} = \cdots = \mu_{1j} = \cdots = \mu_{1n}, \mu_{21} = \cdots = \mu_{i1} = \cdots = \mu_{m1}$. We will show for this maximal entry that $\limsup \mu_{11} < \infty$ if and only if $\liminf[(1 + n/r_1)(1 + m/c_1)/(1 + nm/T)] > 1$.

(34)
$$
\begin{aligned}
(n-1)\mu_{1n} + \mu_{11} &= r_1, \\
(m-1)\mu_{m1} + \mu_{11} &= c_1, \\
(n-1)\mu_{mn} + \mu_{1n} &= (T - r_1)/(m-1), \\
1 + 1/\mu_{11} &= (1 + 1/\mu_{m1})(1 + 1/\mu_{1n})/(1 + 1/\mu_{mn}).
\end{aligned}
$$

It follows that $\mu_{1n} \approx 1, \mu_{m1} \approx 1, \mu_{mn} \sim T/mn$. If $\limsup \mu_{11} < \infty$, then $\mu_{1n} \sim r_1/n, \mu_{m1} \sim c_1/m$, and

(35)
$$
(1 + 1/\mu_{11}) \sim (1 + n/r_1)(1 + m/c_1)/(1 + nm/T) > 1.
$$

Conversely, if $\liminf[(1 + n/r_1)(1 + m/c_1)/(1 + nm/T)] > 1$,
(36)
$$
(1+1/\mu_{11}) = (1+1/\mu_{m1})(1+1/\mu_{1n})/(1+1/\mu_{mn}) > (1+n/r_1)(1+m/c_1)/(1+1/\mu_{mn}),
$$

so also $\liminf(1 + 1/\mu_{11}) > 1$, which implies $\limsup \mu_{11} < \infty$, as required. This concludes the proof of Lemma 3.1.

Lemma 3.3

(37)
$$
\log |V_d| - d \log d = O(d).
$$

Let $\delta_{ij} = 1$ if $i = j, \delta_{ij} = 0$ if $i \neq j$. Then

(38)
$$
E_d(t_{ir} t_{js}) + O(d^{-2}) \approx [\delta_{ij} + \delta_{rs}] d^{-1}.
$$

Proof:
Let $\lambda_{jk} = \mu_{jk}(1 + \mu_{jk})$. The quadratic form $t'V_d t = K_d^2 = \sum_{jk} t_{jk}^2 \lambda_{jk}$ is increasing in each λ_{jk}, so that the determinant $|V_d|$ is also increasing in each λ_{jk}; thus $|V_d| \leq |V_d(\lambda_{11})|$ where $V_d(\lambda_{11})$ is the covariance matrix corresponding to the quadratic form $K_d^2 = \sum_{jk} t_{jk}^2 \lambda_{11}$, for which $|V_d(\lambda_{11})| = \lambda_{11}^d m^{n-1} n^{m-1}$. Similarly, $|V_d| \geq |V_d(\lambda_{mn})| = \lambda_{mn}^d m^{n-1} n^{m-1}$. Thus $\log |V_d| - d \log d = O(d)$. This result may also be obtained by noting that $|V_d|$ is a sum of $m^{n-1} n^{m-1}$ products of d coefficients λ_{jk}.

Again, since the quadratic form $t'V_d t$ is increasing in each λ_{jk}, necessarily the quadratic form $t'V_d^{-1} t$ is decreasing in each λ_{jk}, so bounds for the variances induced

by $t \sim N(0, V_d^{-1})$ are obtained by setting all the λ_{jk} equal to λ_{11} or to λ_{mn}. This establishes that $E_d t_i^2 \approx d^{-1}$.

To bound the off-diagonal terms in V_d, note that $t \sim N(0, V_d^{-1})$ allows us to determine the conditional distribution $v|w$ from the quadratic form $t'V^{-1}t$ with w fixed, and similarly the conditional distribution $w|v$. Indeed the v_j are independent given w, and the w_k are independent given v. This gives a relationship between the v and w covariance matrices which produces the required bound on the off-diagonal terms. A result similar to lemma 3.3 is proved in [BH09] using non-probabilistic methods.

(39)
$$\text{Define } \alpha = 1/\sum_{ij} \lambda_{ij}, \quad \alpha_i = 1/\sum_j \lambda_{ij}, \quad \alpha_{ij} = \alpha_i \lambda_{ij},$$
$$\bar{v} = \alpha \sum_{ij} \lambda_{ij} v_i, \quad \bar{w} = \alpha \sum_{ij} \lambda_{ij} w_j, \quad \tilde{v}_i = v_i - \bar{v}, \quad \tilde{w}_j = w_j - \bar{w}.$$

Note that $\mu_{11} \approx \mu_{mn} \approx 1 \Rightarrow \min_{ij}(\alpha_{ij}/\alpha \sum_k \lambda_{kj}) - \varepsilon \geq 0$ some $\varepsilon > 0$.
And $v_i|w \sim N(-\sum_r \alpha_{ij} w_j, \alpha_i)$ independently for different i. Then

(40)
$$\begin{array}{rcl}
E\{\bar{v}|w\} & = & E\{\sum_i (\alpha/\alpha_i) v_i | w\} = -\sum_i \alpha \sum_r \lambda_{ij} w_j = -\bar{w}, \\
E\{\tilde{v}_i|w\} & = & -\sum_j \alpha_{ij} \tilde{w}_j, \\
E_d\{v_i v_j|w\} & = & \alpha_i \delta_{ij} + \sum_{rs} \alpha_{ir} \alpha_{js} w_r w_s, \\
E_d\{\tilde{v}_i v_j|w\} & = & \alpha_i \delta_{ij} - \alpha + \sum_{rs} \alpha_{ir} \tilde{w}_r \alpha_{js} w_s, \\
E_d\{\tilde{v}_i \bar{v}|w\} & = & \sum_{rs} \alpha_{ir} \tilde{w}_r \bar{w}, \\
E_d\{\tilde{v}_i \tilde{v}_j\} & = & \alpha_j \delta_{ij} - \alpha + \sum_{rs} \alpha_{ir} \alpha_{js} E_d\{\tilde{w}_r \tilde{w}_s\}, \\
& = & \alpha_j \delta_{ij} - \alpha + \sum_{rs}(\alpha_{ir} - \varepsilon\alpha \sum_k \lambda_{kr})\alpha_{js} E_d\{\tilde{w}_r \tilde{w}_s\} \text{ since } \sum_{kr} \lambda_{kr} \tilde{w}_r = 0.
\end{array}$$

Note that $E_d t_i^2 \approx d^{-1} \Rightarrow E_d \tilde{w}_r^2 = O(d^{-1})$.
Also $\alpha \approx d^{-2}, \max \alpha_{ij} \approx d^{-1}, \tilde{\alpha}_{ir} = \alpha_{ir} - \varepsilon\alpha \sum_k \lambda_{kr} \geq 0, \sum_r \tilde{\alpha}_{ir} = 1 - \varepsilon$.

(41)
$$\begin{array}{rcl}
E_d \tilde{v}_i \tilde{v}_j & \leq & \alpha_j \delta_{ij} - \alpha + O(d^{-1}) \max(E_d \tilde{w}_r^2) + \sum_{r \neq s} \tilde{\alpha}_{ir} \alpha_{js} \max_{r \neq s} E_d \tilde{w}_r \tilde{w}_s, \\
& \leq & \alpha_j \delta_{ij} + O(d^{-2}) + (1 - \varepsilon) \max_{r \neq s} E_d \tilde{w}_r \tilde{w}_s, \\
\max_{i \neq j} E_d \tilde{v}_i \tilde{v}_j & \leq & O(d^{-2}) + (1 - \varepsilon) \max_{r \neq s} E_d \tilde{w}_r \tilde{w}_s.
\end{array}$$

Similarly,

(42)
$$\begin{array}{rcl}
\min_{i \neq j} E_d \tilde{v}_i \tilde{v}_j & \geq & O(d^{-2}) + (1 - \varepsilon) \min_{r \neq s} E_d \tilde{w}_r \tilde{w}_s, \\
\max_{i \neq j} |E_d \tilde{v}_i \tilde{v}_j| & \leq & O(d^{-2}) + (1 - \varepsilon) \max_{r,s} |E_d \tilde{w}_r \tilde{w}_s|.
\end{array}$$

The joint distribution of the $\tilde{v}_i, \tilde{w}_r$ depends on the joint distribution of the t_{jk} and so does not depend on the particular particular linear combination of $v_j, 1 \leq j \leq m, w_k, 1 \leq k \leq n$ that is set zero to reduce the dimensionality of these $m + n$ terms to $d = (m + n - 1)$. Thus the reverse result holds conditioning on the $\tilde{v}_i$:

(43)
$$\begin{array}{rcl}
\max_{r \neq s} |E_d \tilde{w}_r \tilde{w}_s| & \leq & O(d^{-2}) + (1 - \varepsilon) \max_{i \neq j} |E_d \tilde{v}_i \tilde{v}_j| \\
\max_{i \neq j} |E_d \tilde{v}_i \tilde{v}_j|, \max_{r \neq s} |E_d \tilde{w}_r \tilde{w}_s| & = & O(d^{-2})
\end{array}$$

A similar argument shows that $\max_{i,j} |E_d \tilde{w}_i \tilde{v}_j| = O(d^{-2})$. Also

(44)
$$t'Vt = \sum_{ij} t_{ij}^2 \lambda_{ij} = \sum_{ij} (\tilde{v}_i + \tilde{w}_j)^2 \lambda_{ij} + (\bar{v} + \bar{w})^2/\alpha$$

so that $\bar{v} + \bar{w}$ is independent of $\tilde{v}_i, \tilde{w}_j$ with variance $\alpha \approx d^{-2}$. Concluding the proof of lemma 3.3,

(45)
$$
\begin{aligned}
E_d(t_{ir}t_{js}) &= E_d(\tilde{v}_i + \tilde{w}_r + \bar{v} + \bar{w})(\tilde{v}_j + \tilde{w}_s + \bar{v} + \bar{w}) \\
&= E_d\tilde{v}_i\tilde{v}_j + E_d\tilde{v}_i\tilde{w}_s + E_d\tilde{w}_r\tilde{v}_j + E_d\tilde{w}_r\tilde{w}_s + E_d(\bar{v} + \bar{w})^2 \\
E_d(t_{ir}t_{js}) + O(d^{-2}) &\approx [\delta_{ij} + \delta_{rs}]d^{-1}.
\end{aligned}
$$

We now apply theorem 1 by verifying the conditions I-V. Similar propositions to I-IV are proved using similar methods in [BH09].

CONDITION I: $\kappa_d^3 = O(1), \kappa_d^4 = O(1)$.

(46)
$$
\begin{aligned}
\kappa_d^3 = E_d(K_d^3)^2 &= E_d(\sum_{jk} t_{jk}^3 \mu_{jk}(1+\mu_{jk})(1+2\mu_{jk}))^2 = O(\sum_{jkrs}|E_d t_{jk}^3 t_{rs}^3|), \\
E_d t_{jk}^3 t_{rs}^3 &= 9 E_d t_{jk}^2 E_d t_{rs}^2 E_d t_{jk} t_{rs} + 6(E_d t_{jk} t_{rs})^3
\end{aligned}
$$

From lemma 3.3,

(47)
$$
\begin{aligned}
E_d t_{jk} t_{rs} &= O(d^{-2} + (\delta_{jr} + \delta_{ks})d^{-1}), \\
E_d t_{jk}^3 t_{rs}^3 &= O(d^{-4} + (\delta_{jr} + \delta_{ks})d^{-3}).
\end{aligned}
$$

In the $O(d^4)$ terms in the sum $\sum_{jkrs} E_d t_{jk}^3 t_{rs}^3$, there are $O(d^3)$ terms in which $(\delta_{jr} + \delta_{ks}) > 0$; thus the sum over all terms is $O(1)$.

$\kappa_d^4 = E_d K_d^4$ is the sum of d^2 terms of $O(d^{-2})$, so it also is bounded.

CONDITION II: $E_d\{Q_\varepsilon \exp[\frac{1}{12}K_d^4(t)]\} = O(1)$.
For X, Y joint normal with mean zero,

(48)
$$
\begin{aligned}
\mathrm{cov}(X^4, Y^4) &= 72 EX^2 EY^2 E^2 XY + 24 E^4 XY \\
\mathrm{cov}(t_{jk}^4, t_{rs}^4) &= O(d^{-6} + (\delta_{jr} + \delta_{ks})d^{-4})
\end{aligned}
$$

Since there are only d^3 covariances for which $(\delta_{jr} + \delta_{ks}) > 0$,

(49)
$$
E_d(K_d^4 - \kappa_d^4)^2 = E_d(K_d^4 - E_d K_d^4)^2 = O(\sum_{jkrs}|\mathrm{cov}(t_{jk}^4, t_{rs}^4)|) = O(d^{-1}).
$$

From [D87] Corollary 5, since $K_d^4 - \kappa_d^4$ is a polynomial of degree 4 in Gaussian variables,

(50)
$$
\begin{aligned}
r > 1 \Rightarrow E_d|K_d^4 - \kappa_d^4|^{2r} &\leq r^{4r}[E_d(K_d^4 - \kappa_d^4)]^r \leq C_r d^{-r}, \\
P_d\{K_d^4 \geq \kappa_d^4 + 1\} &\leq C_r d^{-r}.
\end{aligned}
$$

When $t \sim N(0, V_d^{-1})$, the multivariate normal density is $A\exp[-\frac{1}{2}K_d^2(t)]$. Thus $E_d \exp[\alpha K_d^2(t)] = (1 - 2\alpha)^{-d}$. Also, since the μ_{jk} are bounded, $K_d^4 Q_\varepsilon \leq C\varepsilon^2 K_d^2 Q_\varepsilon$. Thus

(51)
$$
\begin{aligned}
E_d Q_\varepsilon \exp[\frac{1}{12}K_d^4(t)] &\leq E_d \exp[\frac{1}{12}(\kappa_d^4 + 1)] + E_d\{K_d^d \geq \kappa_d^4 + 1\}\exp(\frac{1}{12}C\varepsilon^2 K_d^2) \\
&= O(1) + E_d^{1/2}\{K_4^d > \kappa_d^4 + 1\}\,(1 - \frac{1}{6}CM^2 \log d/d)^{-d/2} \\
&= O(1) + C_r^{1/2}d^{-r/2}d^{\frac{1}{3}CM^2} = O(1) \text{ for } r > CM^2.
\end{aligned}
$$

CONDITION III: $Q_\varepsilon[\log \phi_d(t) - \sum_{r=2}^4 K_d^r(t)\frac{i^r}{r!}] = o(1)$.

For a geometric with mean $\mu \approx 1$, the log centered characteristic function ψ_μ has the standard Taylor series expansion:

$$
\begin{aligned}
\log \psi_\mu(t) &= \sum_{r=2}^4 K^r (it)^r / r! + O(1)|t|^5, \\
(52) \qquad \log \phi_d(t) &= \sum_{jk} \log \psi_{\mu_{jk}}(t_{jk}) = \sum_{r=2}^4 K_d^r i^r / r! + O(1) \sum_{jk} |t_{jk}|^5, \\
\sum_{jk} |t_{jk}|^5 &= O(d^2 \varepsilon^5) = o(1),
\end{aligned}
$$

as required.

CONDITION IV: $E_d Q_\varepsilon \exp[-i K_d^3(t)/6 + \kappa_d^3/72 + K_d^4(t)/24 - \kappa_d^4/24]] \to 1$.
We will first show that $K_d^3 = \sum_{jk} t_{jk}^3 \mu_{jk}(1 + \mu_{jk})(1 + 2\mu_{jk})$ has the same moments in the limit as a normal distribution $N(0, \kappa_d^3)$.

Define $u_\alpha = t_{jk}[\mu_{jk}(1 + \mu_{jk})(1 + 2\mu_{jk})]^{1/3}$ where α ranges over the pairs of indices in $A = \{(j,k), 1 \le j \le m, 1 \le k \le n\}$. Let G be the graph on A with edges $(\alpha, \beta) \in G$ whenever either the first or second index of α, β are the same. In particular, $(\alpha, \alpha) \in G$.

$$
\begin{aligned}
(53) \qquad E_d u_\alpha u_\beta &= O(d^{-2 + G(\alpha,\beta)}), \\
E_d u_\alpha^3 u_\beta^3 &= O(d^{-4 + G(\alpha,\beta)}).
\end{aligned}
$$

Let $\{X_\alpha\}$ denote a multivariate normal with $EX_\alpha = 0, EX_\alpha X_\beta = E_d u_\alpha^3 u_\beta^3$. We will show that $K_d^3 = \sum_\alpha u_\alpha^3$ and $\sum_\alpha X_\alpha$ have moments differing by $O(d^{-1})$. The first two moments are identical, by definition, and the odd moments are zero for both variables. For the $2r$th moment:

$$
\begin{aligned}
(54) \qquad E_d (\sum_\alpha u_\alpha^3)^{2r} &= \sum_\alpha E_d(u_{\alpha_1}^3 u_{\alpha_2}^3 .. u_{\alpha_{2r}}^3) \quad , \\
E(\sum_\alpha X_\alpha)^{2r} &= \sum_\alpha E(X_{\alpha_1} X_{\alpha_2} .. X_{\alpha_{2r}}).
\end{aligned}
$$

The terms $E_d(u_{\alpha_1}^3 u_{\alpha_2}^3 .. u_{\alpha_{2r}}^3)$ tend to be larger when many of the pairs of α_i have edges in G; this size is compensated by the fact that fewer sets of $\alpha_1, \alpha_2, ..\alpha_{2r}$ have many edged pairs. In order to count such sets, for each $\alpha_1, \alpha_2, ..\alpha_{2r}$ we define a set of directed trees $\tau(\alpha) = \{\tau_1(\alpha), \tau_2(\alpha), ..\tau_t(\alpha)\}$ on the α-indices $(1, 2, .., 2r)$.

The tree $\tau_1(\alpha)$ is initialised with root 1; then progress through the α-indices in order, attaching j to k if $(\alpha_j, \alpha_k) \in G$, and k is the smallest index already attached to the tree for which $(\alpha_j, \alpha_k) \in G$. The tree $\tau_i(\alpha)$ is constructed similarly on the set of α-indices not attached to the trees $\{\tau_1(\alpha), \tau_2(\alpha), ..\tau_{i-1}(\alpha)\}$; begin with the root r_i, the lowest α-index not attached to previous trees, and progress through the α-indices in order, attaching j to k if $(\alpha_j, \alpha_k) \in G$, and k is the smallest α-index already attached to the tree $\tau_i(\alpha)$ for which $(\alpha_j, \alpha_k) \in G$.

For a set of trees $\tau = \{\tau_1, \tau_2, ..\tau_t\}$ partitioning the α-indices $\{1, ..2r\}$, the number of $\alpha_1, \alpha_2, ..\alpha_{2r}$ for which $\tau(\alpha) = \tau$ is $O(d^{2r+t})$; to see this, consider the ith tree τ_i which has ,say, α-indices $j_1 = r_i, j_2, ..j_{n_i}$. As $\alpha_{j_1}, \alpha_{j_2}, ..\alpha_{j_{n_i}}$ pass through the $O(d^{2n_i})$ possible values in A^{n_i}, α_{j_1} passes through $O(d^2)$ values, but the remaining α_{j_k} in the tree τ_i each pass through only $O(d)$ values, since each such α_{j_k} is constrained by $(\alpha_{j_k}, \alpha_{j_{k'}}) \in G$ for some fixed $k' < k$. Thus the number of $\alpha_{j_1}, \alpha_{j_2}, ..\alpha_{j_{n_i}}$ with $\tau_i(\alpha) = \tau_i$ is $O(d^{n_i+1})$. Noting that $\sum_i n_i = 2r$, the number of $\alpha_1, \alpha_2, ..\alpha_{2r}$ for which $\tau(\alpha) = \tau$ is the product of these quantities $O(d^{2r+t})$.

For a particular $\alpha_1, \alpha_2, ..\alpha_{2r}$ with trees $\tau(\alpha)$ of sizes $n_1, ..n_t$, Wick's formula for $E(X_{\alpha_1} X_{\alpha_2} .. X_{\alpha_{2r}})$ is the sum over all partitions into r sets of pairs of variables, of the product of the covariances for those variables. The maximal order products

occur when the pairs of variables designated in the partition lie as frequently as possible within one of the trees in $\tau(\alpha)$. Smaller order terms may be ignored because their number is bounded for r fixed. If all the tree sizes n_i are even, the maximal product occurs when each pair of variables designated in the partition has an edge in one of the trees; the covariance for each such variable is $O(d^{-3})$, so the product is $O(d^{-3r})$. If there are s odd terms n_i, there are $s/2$ pairs lying in different trees and having smaller covariances, so

$$(55) \qquad \begin{aligned} E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}}) &= O(d^{-3r-s/2}), \\ \sum_{\alpha|\tau(\alpha)=\tau} E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}}) &= O(d^{-r+t-s/2}). \end{aligned}$$

Now

$$(56) \quad -r+t-s/2 = -\frac{1}{2}\sum n_i + t - s/2 = \frac{1}{2}\sum_{n_i \text{ even}}(2-n_i) + \frac{1}{2}\sum_{n_i \text{ odd}}(1-n_i).$$

Thus

$$(57) \qquad \begin{aligned} \sum_{\alpha|\tau(\alpha)=\tau} E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}}) &= O(1) \text{ if } \max n_i \le 2, \\ \sum_{\alpha|\tau(\alpha)=\tau} E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}}) &= O(d^{-1}) \text{ if } \max n_i > 2. \end{aligned}$$

For a particular $\alpha_1, \alpha_2, ..\alpha_{2r}$ with trees $\tau(\alpha)$ of sizes $n_1, ..n_t$, Wick's formula for $E_d(u^3_{\alpha_1} u^3_{\alpha_2}..u^3_{\alpha_{2r}})$ is the summation over all partitions into $3r$ sets of pairs of variables, of the product of the covariances for those variables. Again, the maximal terms occur when the pairs of variables lie as frequently as possible within the trees of τ. If all the tree sizes are even, the maximal product is $O(d^{-3r})$. If there are s odd tree sizes, there are $s/2$ pairs with smaller covariances, so again

$$(58) \qquad \begin{aligned} \sum_{\alpha|\tau(\alpha)=\tau} E_d(u^3_{\alpha_1} u^3_{\alpha_2}..u^3_{\alpha_{2r}}) &= O(1) \text{ if } \max n_i \le 2, \\ \sum_{\alpha|\tau(\alpha)=\tau} E_d(u^3_{\alpha_1} u^3_{\alpha_2}..u^3_{\alpha_{2r}}) &= O(d^{-1}) \text{ if } \max n_i > 2. \end{aligned}$$

Thus, in equation (54) we need only consider summation over α whose trees have maximal size 2. Let $\tau(2k, 2r-2k)$ denote the trees $\{(1),(2),(3),..(2k)(2k+1, 2k+2)...(2r-1, 2r)\}$. There are $\binom{2r}{2k}$ such τ with $2k$ elements of size 1 and $r-k$ elements of size 2.

Case 1. $\rho(0, 2r)$: All trees of size 2

For example, $\alpha = (11), (12), (22), (23)$ has trees $\{(1,2),(3,4)\}$.

For a particular α with this partition, Wick's formula for $E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}})$ gives a term $E(X_{\alpha_1}X_{\alpha_2})..E(X_{\alpha_{2r-1}}X_{\alpha_{2r}})$ of $O(d^{-3r})$ when the Wick's partition corresponds to $\tau(0, 2r)$, and terms of $O(d^{-3r-1})$ when the Wick's partition includes some terms that are not concordant with $\tau(0, 2r)$.

Also, Wick's formula for $E_d(u^3_{\alpha_1} u^3_{\alpha_2}..u^3_{\alpha_{2r}})$ gives a term $E(u^3_{\alpha_1} u^3_{\alpha_2})..E(u^3_{\alpha_{2r-1}} u^3_{\alpha_{2r}})$ of $O(d^{-3r})$ by summing over the partitions of the $6r$ variables u_{α_i} that conform to $\tau(0, 2r)$; for example, the variables $u_{\alpha_1}, u_{\alpha_1}, u_{\alpha_1}, u_{\alpha_2}, u_{\alpha_2}, u_{\alpha_2}$ will be paired in 15 ways. All other partitions of the $6r$ variables have at least one pairing not conforming with $\tau(0, 2r)$, and the corresponding covariance for that pair is $O(d^{-2})$, so that the contribution of all other partitions is $O(d^{-3r-1})$.

By definition, $E(X_{\alpha_1}X_{\alpha_2})..E(X_{\alpha_{2r-1}}X_{\alpha_{2r}}) = E(u^3_{\alpha_1} u^3_{\alpha_2})..E(u^3_{\alpha_{2r-1}} u^3_{\alpha_{2r}})$. Thus

$$(59) \qquad E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}}) = E(u^3_{\alpha_1} u^3_{\alpha_2}..u^3_{\alpha_{2r-1}} u^3_{\alpha_{2r}}) + O(d^{-3r-1}).$$

Case 2 $\rho(2r, 0)$: All trees of size 1.

For a particular α with this partition, Wick's formula for $E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}})$ sums $E(X_{\alpha_{i_1}}X_{\alpha_{i_2}})..E(X_{\alpha_{i_{2r-1}}}X_{\alpha_{i_{2r}}})$ over all partitions of α into r pairs of variables. Wick's formula for $E_d(u^3_{\alpha_1}u^3_{\alpha_2}..u^3_{\alpha_{2r}})$ consists of a leading term in which, for each i, two of the u_{α_i} are paired; the other terms have at least one u_{α_i} paired with three u_α's that it is unlinked to, and the corresponding covariances have smaller order. The leading term is thus the sum $9^r Eu^2_{\alpha_1}Eu^2_{\alpha_2}...Eu^2_{\alpha_{2r}}E(u_{\alpha_{i_1}}u_{\alpha_{i_2}})..E(u_{\alpha_{i_{2r-1}}}u_{\alpha_{i_{2r}}})$ over all partitions of α into r pairs of variables.

Noting that $E(X_{\alpha_{i_1}}X_{\alpha_{i_2}}) = 9Eu^2_{\alpha_1}Eu^2_{\alpha_2}E(u_{\alpha_1}u_{\alpha_2}) + O(d^{-6})$, obtain that

(60)
$$E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}}) = E(u^3_{\alpha_1}u^3_{\alpha_2}..u^3_{\alpha_{2r-1}}u^3_{\alpha_{2r}}) + O(d^{-4r-1}).$$

Case 3 $\rho(2k, 2r - 2k)$: $2k$ trees of size 1, $r - k$ trees of size 2

For a particular α with this tree, Wick's formula for $E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}})$ has leading product terms in which the partition of the $2r$ terms is such that the terms $X_{\alpha_{2k+1}}X_{\alpha_{2k+2}}..X_{\alpha_{2r}}$ are paired conforming to the last $r - k$ trees of size 2 in $\tau(2k, 2r - 2k)$. Thus

(61)
$$E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}}) = E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2k}})E(X_{\alpha_{2k+1}}X_{\alpha_{2k+2}}..X_{\alpha_{2r}}) + O(d^{-3r-k-1})$$

Similarly, for a particular α with this partition, Wick's formula for $E_d(u^3_{\alpha_1}u^3_{\alpha_2}..u^3_{\alpha_{2r}})$ has leading terms in which the partition of the $6r$ terms is such that the terms $u^3_{\alpha_{2k+1}}u^3_{\alpha_{2k+2}}..u^3_{\alpha_{2r}}$ are paired conforming the last $r - k$ trees of size 2 in $\tau(2k, 2r - 2k)$. Thus

(62)
$$E(u^3_{\alpha_1}u^3_{\alpha_2}..u^3_{\alpha_{2r}}) = E(u^3_{\alpha_1}u^3_{\alpha_2}..u^3_{\alpha_{2k}})E(u^3_{\alpha_{2k+1}}u^3_{\alpha_{2k+2}}..u^3_{\alpha_{2r-1}}u^3_{\alpha_{2r}}) + O(d^{-3r-k-1})$$

From the equivalences in case 1 and case 2,

(63)
$$E(u^3_{\alpha_1}u^3_{\alpha_2}..u^3_{\alpha_{2r}}) = E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}}) + O(d^{-3r-k-1})$$

Since there are $O(d^{3r+k})$ different α with the trees $\tau(2k, 2r - 2k)$,

(64)
$$\sum_{\tau(\alpha)=\tau(2k,2r-2k)} E(u^3_{\alpha_1}u^3_{\alpha_2}..u^3_{\alpha_{2r}}) = \sum_{\tau(\alpha)=\tau(2k,2r-2k)} E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}}) + O(d^{-1})$$

Since this equivalence holds for all partitions with element size at most 2, and the contributions from other partitions are negligible,

(65)
$$\sum E(u^3_{\alpha_1}u^3_{\alpha_2}..u^3_{\alpha_{2r}}) = \sum E(X_{\alpha_1}X_{\alpha_2}..X_{\alpha_{2r}}) + O(d^{-1})$$

as required.

We have shown that $K^3_d = \sum_\alpha u^3_\alpha$ and $\sum_\alpha X_\alpha$ have moments differing by $O(d^{-1})$. Since $\sum_\alpha X_\alpha(\kappa^3_d)^{-1/2} \sim N(0,1)$, and a normal random variable is determined uniquely by its moments, $K^3_d(\kappa^3_d)^{-1/2} \to N(0,1)$ in distribution as $d \to \infty$.

For $Z \sim N(0,1)$, $P\{|Z| > A\} \le \exp(-\frac{1}{2}A^2)$.
Thus $Q_\varepsilon \to 1$ in probability as $d \to \infty$, since for, M large enough
(66)
$$P_d\{Q_\varepsilon = 0\} \le \sum_i P_d\{|t_i| > M\sqrt{\log d/d}\} \le d\exp(-M^2\log d/O(1)) \to 0 \text{ as } d \to \infty.$$

Thus $E_d Q_\varepsilon \exp[-iK_d^3(t)T(\kappa_d^3)^{-\frac{1}{2}}] \to E \exp[iTN(0,1)] = \exp(-\frac{1}{2}T^2)$ uniformly in any finite interval $T^2 \leq A$. Since $|\kappa_d^3| \leq C$, the convergence is uniform in $T^2 \leq \kappa_d^3 A/C$. Now choose $A = C$ to get convergence at $T = (\kappa_d^3)^{-1/2}$:

(67) $$E_d Q_\varepsilon \exp[-\frac{1}{6}iK_d^3(t) + \frac{1}{72}\kappa_d^3] \to 1.$$

Since $K_d^4 - \kappa_d^4 \to 0$ in probability, and using condition IV,

(68) $$\begin{aligned} &|E_d Q_\varepsilon \exp[-\frac{1}{6}iK_d^3(t) + \frac{1}{72}\kappa_d^3](\exp\frac{1}{24}[K_d^4(t) - \kappa_d^4] - 1)| \\ \leq\ & CE_d Q_\varepsilon |\exp\frac{1}{24}[K_d^4 - \kappa_d^4] - 1| \to 0 \end{aligned}$$

Thus, as required, $E_d Q_\varepsilon \exp[-\frac{1}{6}iK_d^3(t) + \frac{1}{72}\kappa_d^3 + \frac{1}{24}K_d^4(t) - \frac{1}{24}\kappa_d^4] \to 1$.

CONDITION V: For some M, $\varepsilon = M\sqrt{\log d/d}$, $\int_{Q_\pi - Q_\varepsilon} |\phi_d| / \int_{Q_\varepsilon} |\phi_d| = o(1)$. A similar result is proved in [BH09] using analytic methods.

Proof: We define a probability P_d on $t_1, ... t_{m+n} = v_1, ..v_m, w_1, ..w_n \in (-\pi, \pi]^{m+n}$ with density proportional to $|\phi_d|$. To prove condition V, we need to show, for some M, $P_d\{\max_i |t_i| \geq \varepsilon | t_{m+n} = 0\} \to 0$ as $d \to \infty$. The method evaluates the conditional probability of large deviations in any single parameter t_i when the rest of the parameters are well behaved.

Since the geometric variable is integer, the geometric characteristic function has period 2π, so individual geometric characteristic functions $\psi_{\mu_{jk}}$ have values near 1 when the argument $v_j + w_k$ has values near 2π or -2π. This will not happen for many pairs v_j, w_k, but is best handled by transformation of each v_j and w_k from $(-\pi, \pi]$ to the unit circle $\{x|e^{ix} = 1\}$:

(69) $$\tilde{v}_j = e^{-iv_j}, \tilde{w}_k = e^{iw_k}, \bar{v} = \frac{1}{m}\sum_j \tilde{v}_j, \bar{w} = \frac{1}{n}\sum_k \tilde{w}_k.$$

Lemma 3.4: With constants $O(1)$ independent of d, j, k,

(70) $$\exp[-|\tilde{v}_j - \tilde{w}_k|^2 O(1)] \leq |\psi_{\mu_{jk}}(v_j + w_k)| \leq \exp[-|\tilde{v}_j - \tilde{w}_k|^2 / O(1)].$$

Proof:
For constants $k(\mu), K(\mu)$, and for all t,
(71)
$$\exp[-|e^{it} - 1|^2 k(\mu)] \leq |\psi_\mu(t)|^2 = \frac{1}{1 + \mu(\mu + 1)|e^{it} - 1|^2} \leq \exp[-|e^{it} - 1|^2 K(\mu)].$$

Also $|e^{i(v_j + w_k)} - 1|^2 = |\tilde{v}_j - \tilde{w}_k|^2$. Since $\mu_{jk} \approx 1$, the lemma is proved.

Lemma 3.5 :

(72) $$\begin{aligned} \text{Define } R^2 &= \sum_{jk} |\tilde{v}_j - \tilde{w}_k|^2. \\ \text{Then, for some } M, P_d\{R > d\varepsilon\} &= \exp[-d/O(1)]. \end{aligned}$$

This lemma guarantees that only t values where most of the $|\tilde{v}_j - \tilde{w}_k|$ are small make significant contributions to the probabilities P_d.

Proof : From (70),

(73) $$\prod_{jk} |\psi_{\mu_{jk}}(v_j + w_k)| \leq \exp[-R^2/O(1)].$$

We have previously used $\int$ to denote integration over the d variables $t_1, ..t_{m+n-1}$, and we will now use $\int_{m+n}$ to denote integration over all variables $t_1, ..t_{m+n}$. From conditions I-IV, theorem 2 implies that $\int Q_\varepsilon \phi_d / \hat{P}\{X_d = 0\} \to 1$, so $\int |\phi_d| \geq |\int Q_\varepsilon \phi_d| = \exp(-\frac{1}{2}d \log d + O(d))$. The integral of $|\phi_d|$ over the first $m + n - 1$ parameters is the same for each choice of t_{m+n}, so the integral over all $m + n$ parameters is $\int_{m+n} |\phi_d| = 2\pi \int |\phi_d|$.

Thus, for M large,

(74)
$$\begin{array}{rcl} \int_{m+n}\{R \geq d\varepsilon\}|\phi_d| & \leq & \int_{m+n} \exp[-d^2\varepsilon^2/O(1)] \leq \exp[-M^2 d \log(d)/O(1) + O(d)], \\ P_d\{R \geq d\varepsilon\} & = & \int_{m+n}\{R \geq d\varepsilon\}|\phi_d|/\int_{m+n}|\phi_d| = \exp[-d/O(1)]. \end{array}$$

Lemma 3.5: For M large enough, $\max_i P_d\{|\tilde{v}_i - \bar{w}| > \varepsilon\} = \exp[-d/O(1)]$.

Proof:.

(75)　　　　From lemma 3.5, for some M, $P_d\{R > d\varepsilon\}$　$\to$　0 as $d \to \infty$,

Define $R_{-i} = \sum_{jk, j \neq i} |\tilde{v}_j - \tilde{w}_k|^2$. Of course $R_{-i} \leq R$. For $i \leq m$,

(76) $R_{-i} \leq d\varepsilon \Rightarrow m\sum_k |\tilde{w}_k - \bar{w}|^2 \leq d^2\varepsilon^2 \Rightarrow \min_k |\tilde{w}_k - \bar{w}| \leq (\frac{m}{n} + \frac{n}{m})^{1/2}\varepsilon = \varepsilon_1.$

By the metric inequality, the interval $I_k = \{\tilde{v} | |\tilde{v} - \tilde{w}_k| \leq \varepsilon_1\}$ on the unit circle, of length at least $2\varepsilon_1$, is such that $|\tilde{v} - \bar{w}| \leq 2\varepsilon_1$ for $\tilde{v} \in I$.
Letting $t_{-i} = \{t_j, j \neq i\}$, note that the conditional density of $t_i | t_{-i}$ is proportional to $\prod_k |\psi_{ik}|$. Then, for t_{-i} satisfying $R_{-i} \leq \varepsilon$, and M_2 chosen large enough,

(77)
$$\begin{array}{rcl} \exp[-\sum_k |\tilde{v}_i - \tilde{w}_k|^2 O(1)]| & \leq & \prod_k |\psi_{ik}| \leq \exp[-d|\tilde{v}_i - \bar{w}|^2/O(1)], \\ P_d\{|\tilde{v}_i - \bar{w}| > \varepsilon_2 | t_{-i}\} & \leq & \exp[-d\varepsilon_2^2/O(1)]/\int \prod_k |\psi_{ik}|dt_i, \\ 1 \geq P_d\{|\tilde{v}_i - \bar{w}| \leq 2\varepsilon_1 | t_{-i}\} & \geq & \exp[-d\varepsilon_1^2 O(1)] \int |\tilde{v}_i - \bar{w}| \leq 2\varepsilon_1\}dt_i/\int \prod_k |\psi_{ik}|dt_i, \\ 1 & \geq & 2\varepsilon_1 \exp[-d\varepsilon_1^2 O(1)]/\int \prod_k |\psi_{ik}|dt_i \\ P_d\{|\tilde{v}_i - \bar{w}| > \varepsilon_2 | t_{-i}\} & \leq & \exp[-d\varepsilon_2^2/O(1) + d\varepsilon_1^2 O(1)]/2\varepsilon_1 \\ & = & \exp[-d/O(1)]. \end{array}$$

The same M_2 holds for all i because $\mu_{jk} \approx 1$, so the $O(1)$ bounds hold for all i. Finally, again with the same $O(1)$ for all i,

(78)
$$\begin{array}{rcl} P_d\{|\tilde{v}_i - \bar{w}| > \varepsilon_2\} & = & P_d\{P_d\{|\tilde{v}_i - \bar{w}| > \varepsilon_2 | t_{-i}\}\{R_{-i} \leq \varepsilon\}\} + P_d\{P_d\{|\tilde{v}_i - \bar{w}| > \varepsilon_2 | t_{-i}\}\{R_{-i} > \varepsilon\}\} \\ & \leq & \exp[-d/O(1)]P_d\{R_{-i} \leq \varepsilon\} + P_d\{R_{-i} > \varepsilon\} \\ \max_i P_d\{|\tilde{v}_i - \bar{w}| > \varepsilon_2\} & = & \exp[-d/O(1)]. \end{array}$$

Now, under P_d, the variable $\tilde{w}_n$ is independent of the variable $\max_{ij} |\tilde{t}_i - \tilde{t}_j|$. Also, if $\max_{ij} |\tilde{t}_i - \tilde{t}_j| \leq \varepsilon \leq 1, \tilde{w}_n = 1$, then $\max_i |t_i| \leq 2\varepsilon$. (We need to constrain ε so that $\max_i |t_i| \leq \pi/2$ to avoid difficulties with the period 2π of the geometric characteristic function.) Then, for some constants M_2, M_3, M_4, M_5, M_6,

$$
\begin{aligned}
P_d\{\max |\tilde{v}_i - \bar{w}| > \varepsilon_2\} \le \sum_i P_d\{|\tilde{v}_i - \bar{w}| > \varepsilon_2\} &= \exp(-d/O(1)) \\
P_d\{\max |\tilde{w}_i - \bar{v}| > \varepsilon_3\} &= \exp(-d/O(1)) \\
P_d\{\bar{w} - \bar{v}| > \varepsilon_4\} &= \exp(-d/O(1)) \\
P_d\{\max_{ij} |\tilde{t}_i - \tilde{t}_j| > \varepsilon_5\} &= \exp(-d/O(1)) \\
P_d\{\max_{ij} |\tilde{t}_i - \tilde{t}_j| > \varepsilon_5|\tilde{w}_n = 1\} &= \exp(-d/O(1)) \\
P_d\{\max_i |t_i| \le \varepsilon_6|w_n = 0\} &= \exp(-d/O(1))
\end{aligned}
\tag{79}
$$

This concludes the proof of the validity of condition V.

4 Equal row and column sums

Consider the special case of [CM07]where the row sums are equal, and the column sums are equal, so that $r_i = \mu n, c_j = \mu m$. In this case $|V_d| = n^{m-1}n^{m-1}\sigma^{2(m+n-1)}$ where $\sigma^2 = \mu(1+\mu)$. In moment calculations, it is convenient to consider the linear transform

$$
(80) \quad
\begin{aligned}
U &= \sum_j v_j/m + \sum_k w_k/n, \\
V_j &= v_j + \sum_k w_k/n, 1 \le j \le m, \\
W_k &= w_k + \sum_j v_j/m, 1 \le k \le n.
\end{aligned}
$$

Note that $Q_\varepsilon(t) = 1 \Rightarrow |U| \le 2\varepsilon, |V_j| \le 2\varepsilon, |W_k| \le 2\varepsilon$. When $t \sim N(0, V_d^{-1})$, the U, V, W are multivariate Gaussian in d dimensions with

$$
(81) \quad
\begin{aligned}
U &\sim N(0, 1/mn\sigma^2), \\
V_j &\sim N(0, 1/n\sigma^2) \text{ independent }, 1 \le j \le m, \\
W_k &\sim N(0, 1/m\sigma^2) \text{ independent }, 1 \le k \le n, \\
&\quad U, V_j - U, W_k - U \text{ independent }.
\end{aligned}
$$

Then

$$
(82) \quad
\begin{aligned}
K_d^2 &= [-mnU^2 + n\sum_j V_j^2 + m\sum_k W_k^2]\sigma^2, \\
K_d^3 &= [-mnU^3 + n\sum_j V_j^3 + m\sum_k W_k^3]\sigma^2(1+2\mu), \\
K_d^4 &= [-mnU^4 + n\sum_n V_j^4 + m\sum_k W_k^4 + 6\sum_j (V_j - U)^2 \sum_k (W_k - U)^2]\sigma^2(1+6\sigma^2),
\end{aligned}
$$

$$
(83) \quad
\begin{aligned}
E_d(K_d^3)^2 &= 3(5(m+n-1)^2 - 4(m-1)(n-1))(1+4\sigma^2)/(mn\sigma^2), \\
E_d K_d^4 &= 3(m+n-1)^2(1+6\sigma^2)/(mn\sigma^2), \\
\hat{P}\{S_d = 0\} &= (2\pi\sigma^2)^{-(m+n-1)/2}m^{(1-n)/2}n^{(1-m)/2} \times \\
&\quad \exp([6(m-1)(n-1) - (m^2+n^2-1)(1+1/\sigma^2)]/12mn).
\end{aligned}
$$

Dropping terms $O(1/d)$, the exponential term is $\exp[\frac{1}{2} - (\frac{m}{n} + \frac{n}{m})(1 + 1/\sigma^2)/12]$. Now the number of points satisfying $R = r, C = c$ is estimated as:

$$
(84) \quad
\hat{Q}(R = r, C = c) = \hat{P}\{R = r, C = c\} \exp(I(P)) = \hat{P}(S_d = 0)[(1+\mu)^{1+\mu}\mu^{-\mu}]^{mn}
$$

Using data from [CM07] , page 5,

**Table 1: Estimated number of contingency tables
with given constant row sums and constant column sums**

Rows	Cols	Summand mean	Exact	Edgeworth	[CM07]1.2
10	10	2	$1.10 \ 10^{59}$	$1.12 \ 10^{59}$	$1.23 \ 10^{59}$
3	3	100/3	$1.33 \ 10^{7}$	$1.23 \ 10^{7}$	$1.68 \ 10^{7}$
3	49	49/3	$1.01 \ 10^{68}$	$4.04 \ 10^{147}$	$1.25 \ 10^{68}$
3	9	11	$2.79 \ 10^{21}$	$2.84 \ 10^{21}$	$3.49 \ 10^{21}$
18	18	13/18	$7.95 \ 10^{127}$	$8.05 \ 10^{127}$	$8.50 \ 10^{127}$
30	30	1/10	$2.23 \ 10^{59}$	$2.23 \ 10^{59}$	$2.32 \ 10^{59}$

The hideously bad approximation at $m = 3, n = 49$, mean $= 49/3$ occurs because the n/m terms in the Edgeworth correction are no longer accurate. (In [CM07], Canfield and MacKay express their approximation as a correction to Good's joint hypergeometric approximation, rather than as a correction to the multivariate Gaussian approximation; this approach produces an estimate that does not involve n/m terms.)

5 The number of graphs with a specified degree sequence

Consider a symmetric table of $0-1$ integers $X_{ij} = X_{ji}, X_{ii} = 0, 1 \leq i \leq n, 1 \leq j \leq n$ with given row sums $D_i = \sum_j X_{ij} = d_i$. The row sums are the degrees of the undirected graph in which $X_{ij} = 1$ corresponds to an edge between nodes i, j. As before we use D_i for a random variable, d_i for a particular value. The random variables $\{D_i\}$ take values on $\{0, 1, ..(n-1)\}^n$. We wish to estimate the number of graphs with the specified degree sequence.

The Edgeworth approximation begins with the maximum entropy distribution on $\{X_{ij}\}$ with expectations $ED_i = d_i$, which consists of independent Bernoullis with expectations μ_{ij}:

(85)
$$P\{X_{ij} = x\} = \mu_{ij}^x (1 - \mu_{ij})^{1-x},$$

where

(86)
$$\log(\mu_{ij}/(1 - \mu_{ij})) = \alpha_i + \alpha_j,$$

and the parameters α_i are chosen so that

(87)
$$ED_i = \sum_j \mu_{ij} = r_i,$$

provided that there exist α that solve these equations. See [BH10b] for conditions on the degree sequences for such α's to exist.

The conditional distribution of $\{X_{ij}\}$ given the degrees $\{d_i\}$ is uniform. The number of graphs with the specified degree sequence is

(88)
$$q(D) = P\{D = d\} \exp[I(P)] = P(D = d)/ \prod_{i<j} (1 - \mu_{ij})^{1-\mu_{ij}} \mu_{ij}^{\mu_{ij}}.$$

The probability $P\{D = d\}$ is estimated by

(89)
$$\hat{P}\{D = d\} = 2(2\pi)^{-n/2} |V_n|^{-1/2} \exp(-\kappa_n^3/72 + \kappa_n^4/24)$$

determined by the first four cumulants of D following the program of section 2.

The reason for the initial factor 2 is that the sum of the degrees is even; the lattice of all possible degree sequences has determinant $\Delta = 2$. The characteristic function over the cube $(-\pi, \pi]^n$ concentrates at $t = 0$ and also at $t = \pi$; the Gaussian formula for the integral near $t = 0$ produces the same value near $t = \pi$, so the total integral is twice the formula for the integral near $t = 0$. For nearly regular graphs, graphs whose degrees are in the ratio $1 + o(n^{-1/2})$, the Edgeworth formula reproduces the asymptotic formula in [MW90].

Each element of D is a sum of independent Bernoullis with expectations $\{\mu_{ij}\}$. The validity of the asymptotic estimate depends on the behaviour of the characteristic function of $D - d$, with parameters $t_j, 1 \leq j \leq n$, setting $t_{jk} = t_j + t_k$,

(90)
$$\phi_n(t) = E\{\exp(it'(D - d))\} = \prod_{j<k} \psi_{\mu_{jk}}(t_{jk}) = \prod_{j<k} e^{-it_{jk}\mu_{jk}}(1 + \mu_{jk}e^{it_{jk}})$$

The cumulants $K_n^r(t)$ of $t'D$ are the sums of the corresponding cumulants of the Bernoullis with expectations μ_{jk} and parameters $t_{jk} = t_j + t_k$,

(91)
$$\begin{array}{rcl}
K_n^2 & = & \sum_{j<k} t_{jk}^2 \mu_{jk}(1 - \mu_{jk}) = t'V_n t, \\
K_n^3 & = & \sum_{j<k} t_{jk}^3 \mu_{jk}(1 - \mu_{jk})(1 - 2\mu_{jk}), \\
K_n^4 & = & \sum_{j<k} t_{jk}^4 \mu_{jk}(1 - \mu_{jk})(1 - 6\mu_{jk}(1 - \mu_{jk})).
\end{array}$$

Then the Edgeworth approximation terms are $\kappa_n^3 = E_n(K_n^3)^2, \kappa_n^4 = E_n(K_n^4)$, where the expectation E_n is under the assumption $t \sim N(0, V^{-1})$. We show in [BH10b] that the formula (88) is valid under similar conditions for the contingency table case, namely that the binomial expectations are relatively bounded as n goes to infinity.

6 Regular graphs

Consider a regular graph, where the degrees all equal to d . Then $\mu = d/(n-1)$; let $v = \mu(1-\mu)$.

(92)
$$
\begin{aligned}
V_n(i,j) &= v(1 + \delta_{ij}(n-2)), \\
|V_n| &= 2(n-1)(n-2)^{n-1}v^n, \\
V_n^{-1}(i,j) &= \frac{-\frac{1}{2(n-1)} + \delta_{ij}}{(n-2)v}, \\
E_n(t_i + t_j)(t_r + t_s) &= \frac{[-\frac{2}{n-1} + \delta_{ir} + \delta_{is} + \delta_{jr} + \delta_{js}]}{(n-2)v}.
\end{aligned}
$$

These expectations may be derived directly, without inverting V, by noting that $t'V_n t \sim \chi_n^2$ has mean n and variance $2n$. The final equation is used in evaluating the third and fourth cumulants, using Wick's formula:

(93) $\qquad EX^4 = 3(EX^2)^2, EX^3Y^3 = 9EX^2EY^2EXY + 6(EXY)^2.$

(94)
$$
\begin{aligned}
\kappa_n^3 &= \mu_n(K_n^3)^2 = 6[(1-4v)^2/v][4(n-2)^2 + 1]/[n(n-1)], \\
\kappa_n^4 &= \mu_n K_n^4 = 6(1/v - 1)(n-2)/(n-1).
\end{aligned}
$$

For n even, the estimated number of regular graphs of degree d is

(95)
$$
\begin{aligned}
\hat{P}\{D = d\}\exp(I(P)) &= \hat{P}\{D = d\}[(1-\mu)^{1-\mu}\mu^\mu]^{-(n-1)/2}, \text{ where} \\
\hat{P}\{D = d\} &= 2(2\pi v)^{-n/2}\left[2(n-1)(n-2)^{n-1}\right]^{-1/2} \times \\
&\quad \exp\left(-\frac{1}{3}\left[(1/v - 4)\frac{(n-2)^2 + 1/4}{n(n-1)} + \frac{1}{4}(1/v - 6)\frac{n-2}{n-1}\right]\right), \\
\text{or } \hat{P}\{D = d\} &= \exp\left(-\frac{n}{2}\log(2\pi vn) + 0.5\log 2 + \frac{5}{6} - \frac{1}{12v} + O(\frac{1}{n})\right).
\end{aligned}
$$

The last formula is identical to the formula given by McKay and Wormald in[WM07]. The previous formula improves the accuracy for modest n by carrying the $n-1$ and $n-2$ terms which give the exact contributions from the third and fourth cumulants. Note that the approximation is symmetric about the degree $d = (n-1)/2, \mu = 1/2$. This is as it should be, since the number of regular graphs with degree d is the same as the number of complementary regular graphs with degree $n - 1 - d$.

The estimated number of graphs is maximized at $\mu = 1/2$, taking the value
$$
(2^{n-2}/\pi n)^{n/2}\exp(1/2)\sqrt{2}.
$$

This can't be too far off, since we get $2^{n(n-1)/2}$ graphs by assigning the $n(n-1)/2$ edges in all possible ways, and we would expect most of the degrees in that population of graphs to be about $d = (n-1)/2$. The other terms in the expression are the Gaussian correction to get the degrees exactly d, and then the Edgeworth correction that identifies a constant ratio departure from the Gaussian formula in the limit.

Table 2: Log number of labelled regular graphs
+ error in Edgeworth approximation

Vertices/Degree	3	4	5	6
8	9.87+.06			
9		13.84+.04		
10	16.23+.10	18.01+.04		
11		22.37+.05		
12	23.17+.14	26.90+.06	28.72+.03	
13		31.58+.08		35.28+.03
14	30.60+.18	36.42+.09	40.18+.04	42.04+.03
15		41.39+.10		48.98+.03
16	38.46+.20	46.49+.11	52.31+.06	56.11+.03
17		51.71+.12		63.41*
18	46.68+.23	57.05+.13	65.04+.08	70.88*

- * numbers are not computed, but estimated from the Edgeworth formula
- The approximation works best when the degree is near half the number of vertices, and gets progressively worse for fixed degree as the number of vertices increases. However, the approximations are not too bad even near the edges; for example the error for 40 vertices and degree 2 is .6 on the log scale, which is about a ratio of 2.

7 Irregular Graphs

Consider now graphs with n_1 vertices of degree d_1, n_2 vertices of degree d_2. The maximum entropy summands are independent Bernoullis on the edges with probabilities

p_{11} for the edges $(i,j), 1 \le i < j \le n_1$,

p_{12} for the edges $(i,j), 1 \le i \le n_1 < j \le n_1 + n_2$,

p_{22} for the edges $(i,j), n_1 < i < j \le n_1 + n_2$.

The maximum entropy choice of the p's is the unique solution , when it exists, to

$$\begin{aligned}
(n_1 - 1)p_{11} + n_2 p_{12} &= d_1, \\
(n_2 - 1)p_{22} + n_1 p_{12} &= d_2, \\
\frac{p_{11}}{1-p_{11}} \frac{p_{22}}{1-p_{22}} &= \left(\frac{p_{12}}{1-p_{12}}\right)^2.
\end{aligned} \tag{96}$$

The Bernoulli variances are $v_{ij} = p_{ij}(1 - p_{ij})$. The random degrees D_i have covariance matrix V :

$$\begin{aligned}
V_{ii} &= (n_1 - 1)v_{11} + n_2 v_{12}, 1 \le i \le n_1, \\
V_{ii} &= (n_2 - 1)v_{22} + n_1 v_{12}, n_1 < i \le n_1 + n_2, \\
V_{ij} &= v_{11}, 1 \le i \ne j \le n_1, \\
V_{ij} &= v_{12}, 1 \le i \le n_1 < j \le n_1 + n_2, \\
V_{ij} &= v_{22}, n_1 < i \ne j \le n_1 + n_2, \\
|V| &= ((n_1 - 2)v_{11} + n_2 v_{12})^{n_1 - 1}((n_2 - 2)v_{22} + n_1 v_{12})^{n_2 - 1} \times \\
&\quad [(2n_1 - 2)v_{11} + n_2 v_{12})((2n_2 - 2)v_{22} + n_1 v_{12}) - n_1 n_2 v_{12}^2].
\end{aligned} \tag{97}$$

In the case where $n_1 = n_2 = n/2, d_2 = n - d_1 - 1, n/4 < d_1 < 3n/4$, then $p_{12} = 1/2, p_{11} = 1 - p_{22} = (d_1 - \frac{1}{4}n)/(\frac{1}{2}n - 1), v_{11} = v_{22}, v_{12} = \frac{1}{4}$, and the covariances of the $t_{ij} = t_i + t_j$ needed for κ_n^3, κ_n^4 are:

$$\begin{aligned}
A &= (\tfrac{1}{2}n - 2)v_{11} + n/8 \\
Q &= ((n - 2)v_{11} + n/8)^2 - (n/8)^2, \\
V_{ii}^{-1} &= 1/A + V_{12}^{-1} \\
V_{ij}^{-1} &= \{n/16 - v_{11}[(n-2)v_{11} + n/8]\}/(AQ), 1 < i < j \le n/2, \\
V_{ij}^{-1} &= -\tfrac{1}{4}/Q, 1 \le i \le n/2, n/2 < j \le n, \\
|V| &= ((\tfrac{1}{2}n - 2)v_{11} + n/8)^{n-2}Q, \\
N_{ij} &= \{1 \le i \le n/2\}\{n/2 < j \le n\} + \{n/2 < i \le n\}\{1 \le j \le n/2\}, \\
E_n t_{ij} t_{kl} &= 4V_{12}^{-1} + (\delta_{ik} + \delta_{il} + \delta_{jk} + \delta_{jl})/A + 4(V_{1n}^{-1} - V_{12}^{-1})\{N_{ik} + N_{il} + N_{jk} + N_{jl}\}
\end{aligned} \tag{98}$$

$$\begin{aligned}
K_n^3 &= v_{11}(1 - 2p_{11}(\textstyle\sum_{1 \le j < k \le n/2} t_{jk}^3 - \sum_{n/2 < j < k \le n} t_{jk}^3) \\
K_n^4 &= v_{11}(1 - 6v_{11})(\textstyle\sum_{1 \le j < k \le n/2} t_{jk}^4 + \sum_{n/2 < j < k \le n} t_{jk}^4) - \tfrac{3}{8}\sum_{1 \le j \le n/2 < k \le n} t_{jk}^4
\end{aligned} \tag{99}$$

The Gaussian approximation:

$$\hat{Q}_G\{D = d\} = 2(p_{11} \log p_{11} + p_{22} \log p_{22})^{-n(n-2)/4}(\log 2)^{-n^2/4}(2\pi)^{-n/2}|V|^{-1/2}. \tag{100}$$

The initial 2 is the determinant of the lattice of possible degree sequences. The second term is the contribution from the Bernoulli probabilities, the exponential value of the entropy. The last term is the Gaussian contribution for the probability that $D = d$. The Edgeworth correction multiplies by the factor

$\exp(-\kappa_n^3/72 + \kappa_n^4/24)$ computed by $\kappa_n^3 = E_n(K_n^3)^2, \kappa_n^4 = E_n K_n^4$ where the expectation is taken under the assumption $t \sim N(0, V^{-1})$.

Table 3: Log number of graphs with irregular degree sequences

Degree Sequence	Exact	Gauss	Edgeworth
44443333	9.59	10.22	9.64
666666555555	28.45	29.03	28.46
77777774444444	24.21	24.83	24.33

The Edgeworth formula is significantly more accurate than the Gaussian formula. The Edgeworth formula is more accurate when the degrees are nearly equal.

References

[B09] A. Barvinok, *Asymptotic estimates for the number of contingency tables, integer flows,and volumes of transportation polytopes*, Int. Math. Res. Notices **2009** (2009), 348–385.

[BH09] A. Barvinok and J.A. Hartigan, *An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums*, arXiv:0910.2477 ,2009

[BH10a] A.Barvinok and J.A. Hartigan, *Maximum entropy Gaussian approximations for the number of integer points and volumes of polytopes*, Advances in Applied Mathematics **45** (2010), 252-289

[BH10b] A.Barvinok and J.A. Hartigan, *The number of graphs and a random graphs with a given degree sequence*, arXiv:1003.0356 ,2010

[CM05] E.R.Canfield and B.D.McKay, *Asymptotic enumeration of dense 0-1 matrices with equal row sums and equal column sums*, Electronic J. Combin., **12** (2005), 29.

[CM07] E.R.Canfield and B.D.McKay, *Asymptotic enumeration of Contingency Tables with Constant Margins* arXiv:0703.600v1 (2007).

[CGM08] E.R. Canfield, C.Greenhill, and B.D.McKay, *Asymptotic enumeration of dense 0-1 matrices with specified line sums*, J. Combin. Th., Ser. A **115** (2008), 32–66.

[DE85] P.Diaconis and B.Efron, *Testing for independence in a two-way table: new interpretations of the chi-square statistic. With discussions and with a reply by the authors*, Ann. Statist. **13** (1985), 845–913.

[D87] J.Duoandikoetxea, *Reverse Hölder Inequalities for Spherical Harmonics*, Proc. Am. Math. Soc. **101** (1987), 487–491.

[G76] I.J.Good, *On the application of symmetric Dirichlet distributions and their mixtures to contingency tables*,Ann.Statist. **4** (1976), 1159–1189.

[GMW06] C.Greenhill, B.D.McKay, and X.Wang, *Asymptotic enumeration of sparse 0-1 matrices with irregular row and column sums*, J. Combinatorial Theory, Ser. A, **113** (2006), 291–324.

[J57] E.T.Jaynes, *Information Theory and Statistical Mechanics, Physical Review* **106** (1957),620–630.

[KT03] S.J.Kathman and G.R.Terrell, *Poisson approximation by constrained exponential tilting*, Statistics & Probability Letters **61** 2003, 83–89

[K06] J.E.Kolassa ,*Series Approximation Methods in Statistics, 3rd Edition* (Lecture Notes in Statistics 88), Springer, New York, 2006.

[MW90] B.D.McKay, and N.C.Wormald, *Asymptotic enumeration by degree sequence of graphs of high degree*, European J. Combin. **11** (1990), 565–580

[MG] B.D.McKay, and C.S.Greenhill, *Asymptotic enumeration of sparse nonnegative integer matrices with specified row and column sums*, Adv. in Appl. Math. **41** (2008), 459–481.